Testing for heteroskedasticity of the residuals in fuzzy rule-based models

José Luis Aznarte M.[1] and José M. Benítez[2]

[1] Centre for Energy and Processes, MINES ParisTech (France)
[2] Dept. of Computer Science and A.I., CITIC-UGR, University of Granada (Spain)

Abstract. In this paper, we propose a new diagnostic checking tool for fuzzy rule-based modelling of time series. Through the study of the residuals in the Lagrange Multiplier testing framework we devise a hypothesis test which allows us to determine if the residual time series is homoscedastic or not, that is, if it has the same variance throughout time. This is another important step towards a statistically sound modelling strategy for fuzzy rule-based models.

1 Introduction

In general, once a model is built and estimated, it has to be evaluated. This is true in the Soft Computing framework as well as in the classical Statistics approach. By evaluating a model we understand to find out if the model satisfies a set of quality criteria that allow us to say if the interesting characteristics of the system under study are actually being captured by it or not.

Notwithstanding, this set of evaluation criteria is heavily dependent on several considerations: the final use that the model is built for, the inner characteristics of the system that are to be captured and whether the emphasis is put on the empirical behaviour of the model or if there are theoretical considerations that are considered to be more important. This is evident when we consider the evaluation means used in the Soft Computing field as opposed to those used in the statistical approach to time series analysis.

In the usually engineering-oriented Soft Computing framework, there has been an overwhelming preeminence of just one evaluation criterion, and this has been the *goodness of fit*. Generally, evaluation of a model consists on computing the prediction (or classification) error produced when it is faced with a previously unseen problem of the same type of the one used to estimate it. This measure, in its different flavours (mean squared error, mean average error and so on) is affected by some inherent limitations: it is not very meaningful for a single model unless compared against other models, and is usually range-dependent, which makes it difficult to compare the same model applied to different problems represented by data sets with different characteristics.

On the other hand, the evaluation in the statistical approach to time series has usually more to do with obtaining an estimate of the probability that the model is effectively capturing the interesting characteristics of the data set, and this is achieved through developing hypothesis tests, also known as misspecification tests.

There is a basic assumption behind modelling: a part of the system under study behaves according to a model but there is another part which cannot be explained by it and is usually considered to be white noise. This is the main idea encoded in the expression of the general model

$$y_t = \mathrm{G}(\mathbf{x}_t; \boldsymbol{\psi}) + \varepsilon_t, \tag{1}$$

and it is also behind some of the diagnostic checking procedures.

It is interesting to obtain a precise knowledge about the series of the residuals, $\{\varepsilon_t\}$, by for example determining if its values are independent and normally distributed. If the residuals were not independent, that would mean that the model is failing to capture an important part of the behaviour of the series, and hence it should be respecified. This can be done through the test presented in [2].

Another desirable property that the model should satisfy refers to the variance of the series $\{\varepsilon_t\}$. If a model is properly capturing the inner behaviour of the series, the residuals should have the same variance at any point of the series. Failing to ensure this implies that the model's precission depends on time, and hence that there are parts of the state-space that are not properly modelled. This will affect very negatively to the performance of the model. Thus this situation should be properly detected so that convenient action for modelling is taken.

The current paper paper addresses the detection of this situation when fuzzy rule-based systems are used to model time series. The chosen procedure is throug the defition of a hypothesis test, which we describe and do a preliminary evaluation.

2 Heteroskedasticity in Time Series Modeling

Ethymologically, heteroskedasticity means differing dispersion or variance. In statistics, a time series is called heteroskedastic if it has different variances throughout the time, and homoskedastic if it shows constant variance in the observable period.

Suppose we have a time series $\{y_t\}_{t=1}^n$ and a vector of time series (explanatory variables) $\{\mathbf{x}_t\}_{t=1}^n$. When considering conditional expectations of y_t given $\mathbf{x}_t$, the time series $\{y_t\}_{t=1}^n$ is said to be heteroscedastic if the conditional variance of y_t given $\mathbf{x}_t$ changes with t. This is also referred as conditional heteroscedasticity to emphasize the fact that it is the series of conditional variance that changes and not the unconditional variance.

A graphical representation might help understand heteroskedasticity The left part of figure 1, (which is adapted from [4]), depicts a classic picture of a homoskedastic situation. We can see a regression line estimated via orthogonal least squares in a simple, bivariate model. The vertical spread of the data around the predicted line appears to be fairly constant as X changes. In contrast, the right part of the figure shows a similar model with heteroskedasticity. The vertical

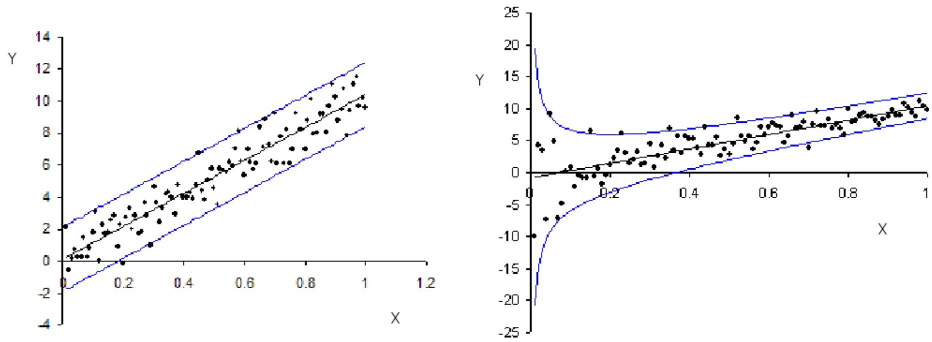

Fig. 1. Example of homoskedastic series (left) and heteroskedastic series (right).

spread of the errors is large for small values of X and then gets smaller as X rises. If the spread of the errors is not constant across the X values, heteroskedasticity is present.

In the case of fuzzy rule-based models for time series analysis, we might be interested in studying the heteroskedasticity of the residual series in the state-space regions defined by the antecedent of the rules. If our model's residual series show smoothly changing variance between the rules, it is likely that some rules are failing to capture the behaviour of the series in their state-space subset. This represents an important source of diagnostic information about the goodness of the model.

3 Fuzzy Rule-based Models for Time Series Analysis

When dealing with time series problems (and, in general, when dealing with any problem for which precision is more important than interpretability), the Takagi-Sugeno-Kang paradigm is preferred over other variants of FRBMs. When applied to model or forecast a univariate time series $\{y_t\}$, the rules of a TSK FRBM are expressed as:

$$\begin{aligned} &\text{IF } y_{t-1} \text{ IS } A_1 \text{ AND } y_{t-2} \text{ IS } A_2 \text{ AND } \ldots \text{ AND } y_{t-p} \text{ IS } A_p \\ &\qquad \text{THEN } y_t = b_0 + b_1 y_{t-1} + b_2 y_{t-2} + \ldots + b_p y_{t-p}. \end{aligned} \quad (2)$$

In this rule, all the variables y_{t-i} are lagged values of the time series, $\{y_t\}$.

Concerning the fuzzy reasoning mechanism for TSK rules, the *firing strength* of the ith rule is obtained as the t-norm (usually, multiplication operator) of the membership values of the premise part terms of the linguistic variables:

$$\omega_i(\mathbf{x}) = \prod_{j=1}^{d} \mu_{A_j^i}(x_j), \quad (3)$$

where the shape of the membership function of the linguistic terms $\mu_{A_j^i}$ can be chosen from a wide range of functions. One of the most common is the Gaussian bell, although it can also be a logistic function and even non-derivable functions as a triangular or trapezoidal function.

The overall output is computed as a weighted average or weighted sum of the rules output. In the case of the weighted sum, the output expression is:

$$y_t = \mathrm{G}\left(\mathbf{x}_t; \boldsymbol{\psi}\right) + \varepsilon_t = \sum_{i=1}^{R} \omega_i(\mathbf{x}_t) \cdot \mathbf{b}_i \mathbf{x}_t + \varepsilon_t, \tag{4}$$

where G is the general nonlinear function with parameters $\boldsymbol{\psi}$, R denotes the number of fuzzy rules included in the system and ε_t is the series of the residuals as mentioned in the Introduction. While many TSK FRBMs perform a weighted average to compute the output, additive FRBMs are also a common choice. They have been used in a large number of applications, for example [5–7, 13].

It has been proved [1] that this specification of the FRBM nests some models from the autoregressive regime switching family. More precisely, it is closely related with the Threshold Autoregressive model (TAR) [12], the Smooth Transition Autoregressive model (STAR) [11], the Linear Local-Global Neural Network (L^2GNN) [10] and the Neuro-Coefficient STAR [9].

This relation has given place to an ongoing exchange of knowledge and methods from the statistical framework to the fuzzy rule-based modelling of time series. For instance, a linearity test against FRBM has been developed [3], and more contributions are yet to come.

In this paper we will consider two types of membership functions: sigmoid, μ_{S}, and Gaussian, μ_{G}. The sigmoid function is the one used in [9], and although it is not so common in the fuzzy literature, we will use it here as an immediate result derived from the equivalences stated in [**?**]. As we know, it is defined as

$$\mu_{\mathsf{S}}(\mathbf{x}_t; \boldsymbol{\psi})) = \frac{1}{1 + \exp\left(-\gamma(\boldsymbol{\omega}\mathbf{x}_t - c)\right)}, \tag{5}$$

where $\boldsymbol{\psi} = (\gamma, \boldsymbol{\omega}, c)$.

On the other hand, Gaussian function will also be used because it is the most common membership function in fuzzy models. It is usually expressed as

$$\mu_{\mathsf{G}}(\mathbf{x}_t; \boldsymbol{\psi})) = \prod_i \exp\left(-\frac{(x_i - c_i)^2}{2\sigma^2}\right) \tag{6}$$

but we will rewrite it as

$$\mu_{\mathsf{G}}(\mathbf{x}_t; \boldsymbol{\psi})) = \prod_i \exp\left(-\gamma(x_i - c_i)^2\right), \tag{7}$$

where $\boldsymbol{\psi} = (\gamma, \mathbf{c})$.

4 Test of homoscedasticity of the residuals of an FRBM

If an FRBM is properly identified and estimated, one might expect that the residuals have a normal distribution, $\varepsilon_t \sim \mathrm{N}(0, \sigma^2)$. Moreover, it is expected that the residuals retain this distribution throughout time, that is, that the mean and the variance of ε_t remain constant through the changes of regime resulting from the prevalence of the different rules in different parts of the state-space.

It is hence interesting to develop a test which can determine if the variance σ^2 of the residual series changes when the model switches from one regime to another or not. Assuming it does vary, we might note it as a time series σ_t^2, whose specification would be:

$$\sigma_t^2 = \sigma^2 + \sum_{i=1}^{r} \sigma_i^2 \mu_{\sigma,i}\left(\mathbf{x}_t; \boldsymbol{\psi}_{\mu_{\sigma,i}}\right) \tag{8}$$

where $\mu_{\sigma,i}$ are sigmoid or Gaussian function satisfying the identifiability restrictions defined in [**?**]. This formulation allows the variance to change smoothly between regimes.

Following [8], in order to avoid complicated restrictions over the parameters to guarantee a positive variance, we rewrite equation (8) as

$$\sigma_t^2 = \exp\left(\mathrm{G}_\sigma\left(\mathbf{x}_t; \boldsymbol{\psi}_\sigma, \boldsymbol{\psi}_{\mu_{\sigma,i}}\right)\right) = \exp\left(\varsigma + \sum_{i=1}^{r} \varsigma_i \mu_{\sigma,i}\left(\mathbf{x}_t; \boldsymbol{\psi}_{\mu_{\sigma,i}}\right)\right), \tag{9}$$

where $\boldsymbol{\psi}_\sigma = [\varsigma, \varsigma_1, ..., \varsigma_r]'$ is a vector of real parameters.

To derive the test, let us consider $r = 1$. This is not a restrictive assumption because the test statistic remains unchanged if $r > 1$. We rewrite model (9) as

$$\sigma_t^2 = \exp\left(\varsigma + \varsigma_1 \mu_\sigma\left(\mathbf{x}_t; \boldsymbol{\psi}_{\mu_\sigma}\right)\right), \tag{10}$$

where μ_σ is defined as (5) or as (7), depending on the membership function used by the model.

In both cases, sigmoid or Gaussian, the null hypothesis of homoscedasticity of the residuals is $\mathrm{H}_0 : \gamma_\sigma = 0$. As usual, model (10) is only identified under the alternative $\gamma_\sigma \neq 0$ and we expand the membership function into a first-order Taylor expansion around $\gamma_\sigma = 0$. Replacing the function by its Taylor approximation and ignoring the remainder, both the sigmoid and the Gaussian case result in

$$\sigma_t^2 = \exp\left(\rho + \sum_{i=1}^{q} \rho_i x_{i,t}\right), \tag{11}$$

so the null hypothesis becomes $\mathrm{H}_0 : \rho_1 = \rho_2 = ... = \rho_q = 0$. Under H_0, $\exp(\rho) = \sigma^2$.

The local approximation to the normal log-likelihood function in a neighbourhood of H_0 for observation t is

$$l_t = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\left(\rho + \sum_{i=1}^{q} \rho_i x_{i,t}\right) - \frac{\varepsilon_t^2}{2\exp(\rho + \sum_{i=1}^{q} \rho_i x_{i,t})}. \tag{12}$$

In order to derive a LM type test, we need the partial derivatives of the log-likelihood:

$$\frac{\partial l_t}{\partial \rho} = -\frac{1}{2} + \frac{\varepsilon_t^2}{2\exp(\rho + \sum_{i=1}^{q} \rho_i x_{i,t})}, \tag{13}$$

$$\frac{\partial l_t}{\partial \rho_i} = -\frac{x_i}{2} + \frac{\varepsilon_t^2 x_i}{2\exp(\rho + \sum_{i=1}^{q} \rho_i x_{i,t})}, \tag{14}$$

and their consistent estimators under the null hypothesis:

$$\left.\frac{\partial \hat{l}_t}{\partial \rho}\right|_{\mathrm{H}_0} = \frac{1}{2}\left(\frac{\varepsilon_t^2}{\hat{\sigma}^2} - 1\right), \tag{15}$$

$$\left.\frac{\partial \hat{l}_t}{\partial \rho_i}\right|_{\mathrm{H}_0} = \frac{x_{i,t}}{2}\left(\frac{\varepsilon_t^2}{\hat{\sigma}^2} - 1\right), \tag{16}$$

where $\hat{\sigma}^2 = 1/T \sum_{t=1}^{T} \hat{\varepsilon}_t^2$. The LM statistic can then be written as

$$LM = \frac{1}{2}\left\{\sum_{t=!}^{T}\left(\frac{\varepsilon_t^2}{\hat{\sigma}^2} - 1\right)\tilde{\mathbf{x}}_t\right\}' \left\{\sum_{t=1}^{T}\tilde{\mathbf{x}}_t\tilde{\mathbf{x}}_t'\right\}^{-1} \left\{\sum_{t=!}^{T}\left(\frac{\varepsilon_t^2}{\hat{\sigma}^2} - 1\right)\tilde{\mathbf{x}}_t\right\} \tag{17}$$

where $\tilde{\mathbf{x}}_t = [1, \mathbf{x}_t]'$. For details, see [8].

The test can be carried out in stages as follows:

1. Estimate model (4) assuming homoscedasticity and compute the residuals $\hat{\varepsilon}_t$. Orthogonalize the residuals by regressing them on $\nabla \mathrm{G}(\mathbf{x}_t; \hat{\psi})$ and as before compute the $SSR_0 = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\tilde{\varepsilon}_t^2}{\hat{\sigma}_{\tilde{\varepsilon}_t}^2} - 1\right)^2$, where $\hat{\sigma}^2$ is the unconditional variance of $\tilde{\varepsilon}_t$.
2. Regress $\left(\frac{\tilde{\varepsilon}_t^2}{\hat{\sigma}_{\tilde{\varepsilon}_t}^2} - 1\right)$ on $\tilde{\mathbf{x}}_t$ and compute the residual sum of squares $SSR_1 = \frac{1}{T}\sum_{t=1}^{T}\tilde{\nu}_t^2$.
3. Compute the χ^2 statistic

$$\mathrm{LM}_{\chi^2}^{\sigma} = T\frac{SSR_0 - SSR_1}{SSR_0}$$

or the F version of the test

$$\mathrm{LM}_F^{\sigma} = \frac{(SSR_0 - SSR_1)}{s}\left(\frac{SSR_1}{(T-s-n)}\right)^{-1}.$$

Where T is the number of observations. Under H_0, $\mathrm{LM}_{\chi^2}^{\sigma}$ is asymptotically distributed as a χ^2 with s degrees of freedom and LM_F^{σ} has approximately an F distribution with s and $T-s-n$ degrees of freedom.

5 Empirical evaluation

In this work we have performed a preliminary assessment of the properties of the test. In this line, we have considered three real-world time series, modeled them with FRMBs and then proceeded to their analysis.

model	#*rules*	sigmoid membership function			Gaussian membership function		
		σ_{ε_t}	AIC	p-value	σ_{ε_t}	AIC	p-value
A	2	0.191	-313	0.179	0.205	-307	0.645
B	2	0.097	-6590	0.000	0.098	-6570	0.000
C	11	0.122	-24357	0.234	0.120	-24516	0.566

Table 1. Results of misspecification tests for three models facing real world cases (significance value: 0.95).

The considered cases are fully described in [1], and are a well known ecology problem (the Lynx series), a planning/management problem and a botanic problem.

The first series, commonly referred to as the Lynx series, is composed of the annual records of lynx captures in a certain part of Canada during a period spanning 113 years. It is a common benchmarking series used to test and compare time series models, and here we have used its logarithmic transformation. An FRBM with two rules (model A) was identified following the iterative procedure proposed in [1], and it was later estimated using a Genetic Algorithm.

The second considered series comes from an emergency call center and is the record of the number of calls received daily throughout four years. As the series is non-stationary and shows a high variability, it was differenced after applying a log-transformation. The identified FRBM (model B) was also composed of just two fuzzy rules, which were also fine tuned through a Genetic Algorithm.

Finally, the third series was a daily aerobiological log obtained over sixteen years in the city of Granada (Spain), containing daily counts of airborne olive tree pollen grains. This series was previously studied in [**?**].

Table 1 shows some information about the application of the FRBM, both in its sigmoid and Gaussian versions, to the three time series mentioned above. More precisely, for each model, the table shows, the number of rules of the model, the values for the variance of the residuals (σ_{ε_t}) and the Akaike information criterion (AIC), together with the p-value obtained with the test for homoscedasticity of the residuals.

By studying the p-values shown in columns 5 and 8 we can see how the null hypothesis of the test was rejected in all the six cases, which leads us to conclude that the variance of the residuals remained constant through time in every application.

As mentioned above, this is a necessary condition for considering that a model is properly capturing the behaviour of a time series.

6 Conclusions and Final Remarks

In this paper, a new statistical tool to evaluate the residuals of a fuzzy rule-based model has been presented. It consists of a test against homoskedasticity of the residuals, that is, a test that allow the user to determine if the variance of the residual series remains constant through time. In other words, this test is able to tell if a model's errors are bigger in some parts of the state-space or not.

This represents a useful contribution and another step towards a statistically sound framework for the use of fuzzy rule-based models.

Acknowledgements: This work has been partially funded by Spanish Ministerio de Ciencia e Innovación (MICINN) under Project grants MICINN TIN2009-14575 and CIT-460000-2009-46.

References

1. Aznarte M., J.L.: Modelling time series through fuzzy rule-based models: a statistical approach. Ph.D. thesis, Universidad de Granada (2008)
2. Aznarte M., J.L., Benítez, J.M.: Testing for linear independence of the residuals in the framework of fuzzy rule-based models. In: Ninth International Conference on Intelligent Systems Design and Applications (ISDA). Pisa (Italy) (Dec 2009)
3. Aznarte M., J.L., Medeiros, M., Sánchez, J.M.B.: Linearity testing against a fuzzy rule-based model. Fuzzy Sets and Systems (in press)
4. Barreto, H., Howland, F.: Introductory Econometrics: Using Monte Carlo Simulation with Microsoft Excel. Cambridge University Press (Dec 2005)
5. Byun, H., Lee, K.: A decision support system for the selection of a rapid prototyping process using teh modified topsis method. Intern. Journal of Advanced Manufacturing Technology 26(11–12), 1338–1347 (2005)
6. John, R., Innocent, P.: Modeling uncertainty in clinical diagnosis using fuzzy logic. Systems, Man, and Cybernetics, Part B, IEEE Transactions on 35(6), 1340–1350 (Dec 2005)
7. Lee, I., Kosko, B., Anderson, W.F.: Modeling gunshot bruises in soft body armor with an adaptive fuzzy system. IEEE Transactions on Systems, Man, and Cybernetics, Part B 35(6), 1374–1390 (2005)
8. Medeiros, M., Veiga, A.: Diagnostic checking in a flexible nonlinear time series model. Journal of Time Series Analysis 24, 461–482 (2003)
9. Medeiros, M., Veiga, A.: A flexible coefficient smooth transition time series model. IEEE Transactions on Neural Networks 16(1), 97–113 (January 2005)
10. Suarez-Farinas, M., Pedreira, C.E., Medeiros, M.C.: Local global neural networks: A new approach for nonlinear time series modeling. Journal of the American Statistical Association 99, 1092–1107 (December 2004), http://ideas.repec.org/a/bes/jnlasa/v99y2004p1092-1107.html
11. Teräsvirta, T.: Specification, estimation and evaluation of smooth transition autoregresive models. J. Am. Stat. Assoc. 89, 208–218 (1994)
12. Tong, H.: On a threshold model. Pattern Recognition and Signal Processing (1978)
13. Vieira, C.F., Palma, L.B., da Silva, R.N.: Robust fault diagnosis approach using analytical and knowledge based techniques applied to a water tank system. International journal of engineering intelligent systems for electrical engineering and communications 13(4), 237 – 244 (2005)